Proceedings of the 5th International Conference on Statistics: Theory and Applications (ICSTA'23)
Brunel University, London, United Kingdom – August 03-05, 2023
Paper No. 111
DOI: 10.11159/icsta23.111

Multivariate Permutation Test for Model Selection

Elena Barzizza[1], Riccardo Ceccato[1], Luigi Salmaso[1]
[1]University of Padova, Department of Management Engineering
Stradella San Nicola 3, 36100 Vicenza, Italy
elena.barzizza@phd.unipd; riccardo.ceccato.1@unipd.it; luigi.salmaso@unipd.it

Extended Abstract

In the last two decades, the availability of data from different sources in all industrial fields has grown [1]. Thus, big data analysis and machine learning (ML) techniques have become fundamental in firms' decision-making processes. The possibility to make reliable forecasting is a crucial step for companies that can have a beneficial impact on their profits. In the literature, many ML algorithms are available both for classification and regression tasks. However, the choice of the best-performing algorithm remains a crucial aspect. Hitherto different metrics have been proposed to evaluate the performance of that models and some solutions are provided in literature to face with machine learning model selection problem. This problem of the choice of ML model arises in every approach, the basic idea is to consider a train set and a test set starting from the dataset considered: the first is useful for training the ML model while the second is used to evaluate its performance. Hence, in this context, we aim to introduce an innovative and robust testing procedure based on the NonParametric combination methodology (Pesarin & Salmaso 2010 [3]) to select the best-performing machine learning models in a regression task. This is a permutation-based methodology so it is a distribution-free method, it needs few requirements, and it allows high flexibility on the choice of the test statistic according to the nature of the problem. We would consider four different measures of prediction errors for ML models, namely the absolute error (ae), the squared error (se), the absolute percent error (ape) and the squared log error (sle). As stated in Botchkarev (2019) [2] in literature we find a considerable list of error measures but there are no references on which is the best ever.

Our objective is to rank the C applied ML models from the best one to the worst one in terms of performances according to the aforementioned error measures. Thus, our null hypothesis deals with the equality of cumulative distribution functions against a one-sided ordered alternative with at least a strict inequality. The NPC requires the decomposition of this system in four subsystems, one for each error metric considered. The methodology needs to perform all the possible pairwise comparison among groups, from these comparisons we can derive our partial p-values by applying our NPC-based test statistic for each sub-system. Finally, a combination step is necessary to retrieve our global p-value to take a univocal decision on the ranking among groups. Since we deal with paired samples, we need to adopt an appropriate permutation approach to implicitly consider the dependency.

We apply our robust procedure to a real case study. We will focus on the application of some ML algorithms including, Random Forest (ranger), Stochastic Gradient Boosting (gbm), CART (rpart), Cubist (cubist) and Bayesian Generalized Linear Model (bayesglm). These ML models will be compared using the error measures mentioned above. The results of this application will give interesting insights to practitioners.

References

[1] Addo-Tenkorang, R., & Helo, P. T. (2016). Big data applications in operations/supply-chain management: A literature review. *Computers & Industrial Engineering*, *101*, 528-543.

[2] Botchkarev, A. (2018). Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology. *arXiv preprint arXiv:1809.03006*.
[3] Salmaso, L., & Pesarin, F. (2010). *Permutation tests for complex data: theory, applications and software*. John Wiley & Sons.